

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of

Atty. Docket

RENAT VAFIN ET AL.

NL 010529

Serial No.

Group Art Unit

Filed: CONCURRENTLY

Ex.

Title: CODING SIGNALS

Commissioner for Patents  
Washington, D.C. 20231

AUTHORIZATION PURSUANT TO 37 CFR §1.136(a)(3)  
AND TO CHARGE DEPOSIT ACCOUNT

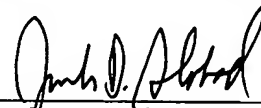
Sir:

The Commissioner is hereby requested and authorized to treat any concurrent or future reply in this application requiring a petition for extension of time for its timely submission, as incorporating a petition for extension of time for the appropriate length of time.

Please charge any additional fees which may now or in the future be required in this application, including extension of time fees, but excluding the issue fee unless explicitly requested to do so, and credit any overpayment, to Deposit Account No. 14-1270.

Respectfully submitted,

By



Jack D. Slobod, Reg. 26,236  
Attorney  
(914) 333-9606

jc997 U.S. PTO  
10/003052  
11/02/01

**THIS PAGE BLANK (USPTO)**



Europäisches  
Patentamt

European  
Patent Office

Office européen  
des brevets

N 000625  
#15 308-03  
Priority papers  
NL010529

Bescheinigung

Certificate

Attestation

jc997 U.S. PTO  
10/003052



Die angehefteten Unterla-  
gen stimmen mit der  
ursprünglich eingereichten  
Fassung der auf dem näch-  
sten Blatt bezeichneten  
europäischen Patentanmel-  
dung überein.

The attached documents  
are exact copies of the  
European patent application  
described on the following  
page, as originally filed.

Les documents fixés à  
cette attestation sont  
conformes à la version  
initialement déposée de  
la demande de brevet  
européen spécifiée à la  
page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

00203857.8

Der Präsident des Europäischen Patentamts;  
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets  
p.o.

I.L.C. HATTEN-HECKMAN

DEN HAAG, DEN  
THE HAGUE,  
LA HAYE, LE

25/05/01

**THIS PAGE BLANK (USPTO)**



Europäisches  
Patentamt

European  
Patent Office

Office européen  
des brevets

**Blatt 2 der Beschreibung  
Sheet 2 of the certificate  
Page 2 de l'attestation**

Anmeldung Nr.:  
Application no.:  
Demande n°: 00203857.8

Anmeldetag:  
Date of filing: 03/11/00  
Date de dépôt:

Anmelder:  
Applicant(s):  
Demandeur(s):  
Koninklijke Philips Electronics N.V.  
5621 BA Eindhoven  
NETHERLANDS

Bezeichnung der Erfindung:  
Title of the invention:  
Titre de l'invention:  
Modifying transients for efficient coding of audio

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat:	Tag:	Aktenzeichen:
State:	Date:	File no.
Pays:	Date:	Numéro de dépôt:

Internationale Patentklassifikation:  
International Patent classification:  
Classification internationale des brevets:

/

Am Anmeldetag benannte Vertragsstaaten:  
Contracting states designated at date of filing: AT/BE/CH/CY/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE/TR  
Etats contractants désignés lors du dépôt:

Bemerkungen:  
Remarks:  
Remarques:

**THIS PAGE BLANK (USPTO)**

# Modifying Transients for Efficient Coding of Audio

Renat Vafin, Richard Heusdens, W. Bastiaan Kleijn

## Summary

We propose an algorithm for efficient representation of transients in audio signals. We estimate the transient part in an original audio signal and modify the locations of the transients in such a way that the transients can occur only at locations specified by a time grid. This procedure allows an efficient representation of transients with damped sinusoids. We also verify that the introduced modifications do not result in a perceptual difference between the original and modified audio signal.

## 1 Introduction

Parametric coding of audio is a popular tool for representing audio signals at very low bit rates [1, 2, 3, 4]. In a parametric audio coder a signal is represented with a source model, and parameters of the model are estimated and encoded. Psychoacoustic rules are used to control the estimation and encoding process in order to exploit irrelevancy present in the signal.

Different source models can be appropriate for different audio signals. The ultimate goal is to obtain an efficient representation of an audio signal while meeting perceptual quality requirements. A popular representation of audio signals is based on decomposition of an original signal into three parts: a transient part, a sinusoidal part and a noise part. Each part is then represented by a corresponding set of parameters [1, 3, 4]. Having a dedicated model for the transient part proved to be beneficial for portions of audio signals with sharp attacks, because sinusoidal and noise model cannot represent such events efficiently [5].

We present a method that improves the efficiency of representing the transient part of an audio signal. It is shown in [6] that transients can be efficiently modeled using exponentially-modulated sinusoids (below we refer to exponentially-modulated sinusoids as damped sinusoids, assuming decaying amplitudes). In that work, an audio signal is analyzed on a segment-by-segment basis, and each segment is represented as a sum of damped sinusoids. A problem occurs when a transient occurs in the middle of a segment. Compared to the case when a transient occurs in the beginning of a segment the number of damped sinusoids needed to model the transient well increases considerably. A possible solution to this problem is to allow a variable segmentation of the signal, such that transients will always occur in the beginning of the segments. However, the search for the optimal segmentation introduces an additional computational cost. Furthermore, additional side information is required to describe the segmentation.

In our new approach we use a fixed segmentation and modify the transient part of the audio signal instead. The modified transient signal can be efficiently represented by damped sinusoids. Specifically, we proceed in the following steps:

1. Estimate the transient part in the original audio signal and subtract it from the original audio signal to form a residual signal. (Next, the residual signal can be analyzed by sinusoidal and noise models).
2. Modify the locations of the estimated transients in such a way that the transients can only occur at locations specified by a time grid.
3. Model the modified transient signal with damped sinusoids.

We also verify that when the modified transient signal is added to the residual signal obtained in step 1, there is no perceptual difference between the resulting signal and the original audio signal. Thus, we can claim that the introduced difference in transients locations is not of perceptual importance, i.e. irrelevant. Thus, no additional side information cost is introduced by the modification.

In order to modify transients locations we have first to estimate the transient part in the original audio signal. Different transient models have been used in parametric coding of audio [1, 2, 7]. For our purpose, a transient model should provide a good approximation of the transient part, and have a lower complexity compared to the damped sinusoids model with variable





Figure 1: Positions of sinusoids in the frequency-domain segment.

segmentation mentioned above. In our current work, we use the elegant transient model based on duality between time and frequency domain presented in [7]. Our transient modification approach based on this model is described in the next section.

## 2 Modification of transients locations

### 2.1 Transient estimation and modification

The transient estimation model presented in [7] is based on the duality between the time and frequency domain. A delta-impulse in the time domain corresponds to a sinusoid in the frequency domain. Furthermore, a sharp transient in the time domain corresponds to a frequency-domain signal which can be represented efficiently by a sum of sinusoids. Specifically, the transients are estimated in the following steps:

1. The discrete cosine transform (DCT) is used to transform a signal to the frequency domain. The DCT size should be sufficiently large to ensure that a transient is a short event in the time-domain segment (thus, its transform to the frequency domain can be modeled efficiently by sinusoids). A block size of about 1 s is sufficient.
2. The frequency-domain (DCT-domain) signal is analyzed with a sinusoidal model. An iterative matching pursuit algorithm is used for the analysis. In our current work we use a consistent iterative sinusoidal analysis-synthesis with Hanning-windowed sinusoids as presented in [8]. Positions of sinusoids in the frequency-domain segment are illustrated in figure 1.

As a result of the above procedure a transient is represented by sinusoidal parameters (amplitudes, frequencies, phases) obtained in step 2. The

information about the location of a transient in a time-domain segment is contained in the frequency parameters of the sinusoids corresponding to the transient. A transient in the beginning of a segment results in low sinusoidal frequencies, while a transient in the end of a segment results in high sinusoidal frequencies. The frequency resolution of the sinusoidal model depends on the required resolution in estimation transients locations. If the required time resolution is one sample then the required frequency resolution is defined by the reciprocal of the DCT size.

Due to the duality between the transient location in a time-domain segment and the frequencies of the corresponding sinusoids, the obvious way to modify the transient location is to modify the corresponding frequencies (plus a correction in the phase parameters). Let us denote by  $n_0$  the transient location in the time-domain segment and  $\{A_{i,j}, \omega_{i,j}, \phi_{i,j}\}$  the sinusoidal parameters corresponding to the transient. Here the index  $i$  is the sinusoid index in a sinusoidal segment and the index  $j$  is the sinusoidal segment index in a DCT-domain segment (both indices have the first value 1). Then in order to modify the transient location by  $\Delta n$  (a positive  $\Delta n$  corresponds to a shift backward in time and a negative  $\Delta n$  to a shift forward in time), the frequencies  $\omega_{i,j}$  and phases  $\phi_{i,j}$  should be modified as follows:

$$\hat{\omega}_{i,j} = \omega_{i,j} - \frac{\Delta n \pi}{N}, \quad (1)$$

$$\hat{\phi}_{i,j} = \phi_{i,j} + \frac{\Delta n \pi}{N} \left( \frac{L-1}{2} + (j-1) \frac{L}{2} \right), \quad (2)$$

where  $N$  is the DCT size,  $L$  is the length of the sinusoidal segments used in the analysis of the DCT-domain segment and  $L/2$  is the shift between the sinusoidal segments. In the sinusoidal segments the phase 0 is defined to be in the middle of the segments. No modification of amplitudes  $A_{i,j}$  is needed.

Note that the above procedure is different from independent quantization of sinusoidal parameters corresponding to the transient. All frequencies are modified by the same amount. This together with the phase correction of equation (2) ensures that the shape of the time-domain transient is preserved, only the location is modified. The shift  $\Delta n$  is defined by the difference between the original location  $n_0$  and the closest allowed location from a time grid.

## 2.2 Details of the modification procedure

- Because the DCT size is large (1 s), more than one transient can occur in a time-domain segment. In this case, the model has to identify sinusoidal parameters corresponding to different transients. This is done by declaring close sinusoidal frequencies  $\omega_{i,j}$  to represent the same transient. Specifically, two sinusoids having frequencies differing by not more than  $\varepsilon_\omega$  are declared to represent the same transient and two sinusoids having frequencies differing by more than  $\varepsilon_\omega$  are declared to represent different transients. Then locations of all transients are modified separately.
- A transient can occur at the very beginning or at the very end of a time-domain segment. Then the modification of sinusoidal frequencies can yield frequencies below 0 or above  $\pi$ . This results in a distortion of the shape of the time-domain transient. Simply cancelling such sinusoids will not preserve the transient shape, because some information describing the transient is not used.

Instead, we use a solution that preserves the transients shapes. We allow an overlap between time-domain segments (about 0.1 ms). In this case a transient can appear in two overlapping segments, i.e. in the region of mutual overlap. Because the overlap is sufficiently large, if the transient is located very close to a border of one of the overlapping segments, then it is located on a safe distance from a border of the other segment. It is straightforward to identify the transient location from sinusoidal frequencies, and therefore it is easy knowing the estimated sinusoidal frequencies in the two overlapping segments to identify when a transient is represented in two segments. If such a situation occurs we cancel the corresponding sinusoids in one of the two segments, in the one where the transient is closer to a border.

- A typical transient lasts for more than one time sample. A natural question is then what is the location  $n_0$  of the transient. In our current approach, we identify the set of frequencies  $\omega_{i,j}$  representing the transient and then define the most common frequency value  $\omega_0$  in the set. The location of the transient is then found from  $\omega_0$  using the duality between the time domain and frequency domain. Currently, we are working on improving the estimation of  $n_0$ .

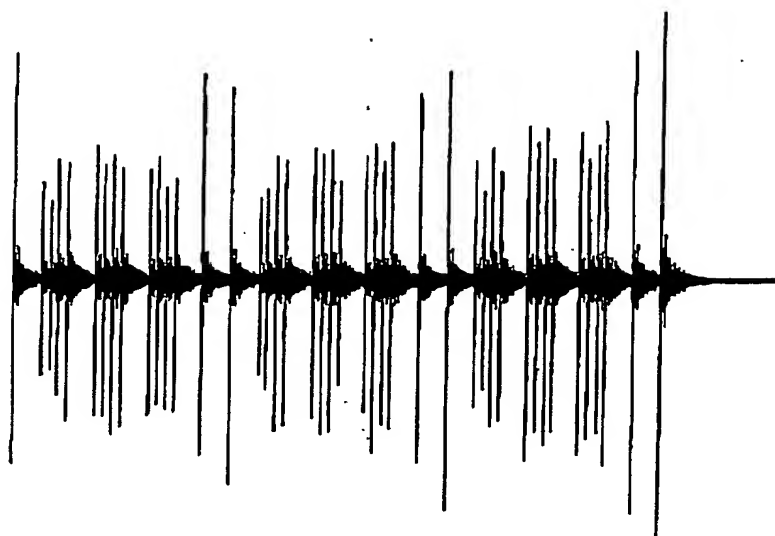


Figure 2: Castanet signal.

## 2.3 Results

Our preliminary experiments were performed on the castanet signal (figure 2). This signal contains short transients and they are well described by the transient estimation technique of [7]. The transient signal was estimated, and the residual signal was obtained as the difference between the original audio signal and the transient signal. The transient signal was modified such that the transients were allowed to occur at locations corresponding to a time grid of 5 ms or 10 ms. In both cases, the modifications did not result in the perceptual difference between the original and modified transient signals. However, when a modified transient signal was added to the residual signal, a slight perceptual difference with the original signal was noticed. We could hear a distortion before transients. We believe that the problem comes from the sinusoidal analysis of the DCT-domain signals, and that the incorporation of the tonality measure defined in [9] and used in [7] will solve the problem. The standard [9] is requested from Philips.

### 3 Ongoing work

- Defining  $n_0$  as described in subsection 2.2.
- Modeling the modified transient signal with damped sinusoids and comparing efficiency of the transient representation with the reference method of [7]. The modeling with damped sinusoids can be performed, e.g. as described in [6].

### 4 Application areas

Audio coding.

### References

- [1] K. N. Hamdy, M. Ali, and A. H. Tewfik. Low bit rate high quality audio coding with combined harmonic and wavelet representation. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, volume 2, pages 1045–1048, Atlanta, Georgia, USA, 1996.
- [2] B. Edler, H. Purnhagen, and C. Ferekidis. ASAC - analysis/synthesis audio codec for very low bit rates. Preprint 4179 (F-6) 100th AES Convention, Copenhagen, 11-14 May 1996.
- [3] H. Purnhagen. Advances in parametric audio coding. In *Proc. 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages W99–1–W99–4, New Paltz, New York, USA, 1999.
- [4] T. S. Verma and T. H. Y. Meng. A 6 kbps to 85 kbps scalable audio coder. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, volume II, pages 877–880, Istanbul, Turkey, 2000.
- [5] M. Goodwin. Residual modeling in music analysis-synthesis. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, volume 2, pages 1005–1008, Atlanta, Georgia, USA, 1996.
- [6] J. Nieuwenhuijse, R. Heusdens, and E. F. Deprettere. Robust exponential modeling of audio signals. In *Proc. IEEE Int. Conf. Acoust., Speech,*

*and Signal Proc.*, volume 6, pages 3581–3584, Seattle, Washington, USA, 1998.

- [7] T. S. Verma, S. N. Levine, and T. H. Y. Meng. Transient modeling synthesis: a flexible analysis/synthesis tool for transient signals. In *Proc. Int. Computer Music Conf.*, pages 25–30, Thessaloniki, Greece, 1997.
- [8] K. Vos, R. Vafin, R. Heusdens, and W. B. Kleijn. High-quality consistent analysis-synthesis in sinusoidal coding. In *Proc. Audio Eng. Soc. 17th Conference "High Quality Audio Coding"*, pages 244–250, Florence, Italy, 1999.
- [9] ISO/MPEG Committee. *Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 Mbit/s - part 3: Audio*.

03-11-2000

11. 2000

PHILIPS 000N EP00203857-8

SPEC

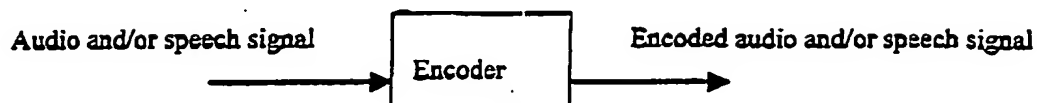


Fig. 3

Fig. 3 shows an audio and/or speech encoder according to an embodiment of the invention. This encoder estimates a transient in the signal, modifies a location of the estimated transient to obtain a modified transient having a location closer to a grid location than the estimated transient location, the grid location being specified by a predefined time grid, and includes a representation of the modified transient in the encoded signal.

CLAIMS:

1. A method of encoding a signal to obtain an encoded signal, the method comprising the steps of:
  - estimating a transient in the signal,
  - modifying a location of the estimated transient to obtain a modified transient having a location closer to a grid location than the estimated transient location, the grid location being specified by a predefined time grid, and
  - including a representation of the modified transient in the encoded signal.
2. An encoder for encoding a signal to obtain an encoded signal, the encoder comprising:
  - means for estimating a transient in the signal,
  - means for modifying a location of the estimated transient to obtain a modified transient having a location closer to a grid location than the estimated transient location, the grid location being specified by a predefined time grid, and
  - means for including a representation of the modified transient in the encoded signal.
3. An encoded signal comprising representations of modified transients having a location approximating a grid location, the grid location being specified by a predefined time grid.
4. A storage medium on which a signal as claimed in claim 3 has been stored.